

# The entropy of network ensembles

Ginestra Bianconi

*The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*

In this paper we generalize the concept of random networks to describe networks with non trivial features by a statistical mechanics approach. This framework is able to describe ensembles of undirected, directed as well as weighted networks. These networks might have not trivial community structure or, in the case of networks embedded in a given space, non trivial distance dependence of the link probability. These ensembles are characterized by their entropy which evaluate the cardinality of networks in the ensemble. The general framework we present in this paper is able to describe microcanonical ensemble of networks as well as canonical or hidden variables network ensemble with significant implication for the formulation of network constructing algorithms. Moreover in the paper we define and characterize in particular the *structural entropy*, i.e. the entropy of the ensembles of undirected uncorrelated simple networks with given degree sequence. We discuss the apparent paradox that scale-free degree distribution are characterized by having small structural entropy but are so widely encountered in natural, social and technological complex systems. We give the proof that while scale-free networks ensembles have small structural entropy, they also correspond to the most likely degree distribution with the corresponding value of the structural entropy.

PACS numbers: 89.75-k, 89.75.Fb, 89.75.Hc

## INTRODUCTION

The quantitative measure of the order present in complex systems and the possibility to extract information from the complex of interactions in cellular, technological and social networks is a topic of key interest in modern statistical mechanics. The field of complex networks [1, 2] has having a rapid development and a large success in this respect due to the wide applicability of simple concepts coming from graph theory. The characterization of the structure of different networks has allowed the scientific community to compare systems of very different nature. Different statistical mechanics tools have been devised to describe the different level of organization of real networks. A description of the structure of a complex network is presently performed by measuring different quantities as (i) the density of the links, (ii) the degree sequence [3], (iii) the degree-degree correlations [4, 5, 6], (iv) the clustering coefficient [7, 8], (v) the k-core structure [9, 10, 11], (vi) the community structure [2, 12, 13, 14] and finally the nature of the embedding space [15, 16, 17]. Moreover, if the network is weighted, strength/degree correlations [18] and if the network is directed, in-degree/out-degree correlations [19] are significant characteristics of the network. These phenomenological quantities describe the local or non-local topology of the network and do affect dynamical models defined of them [1].

While many different statistical mechanics models have been proposed [20, 21, 22, 23, 24, 25, 26, 27] to describe how the power-law degree distribution can arise in complex networks, little work has been done on the problem of measuring the level of organization and "order" in the frame of theoretical statistical mechanics. Only recently, in the field of complex networks atten-

tion has been addressed to the study of entropy measures [28, 29, 30, 31, 32, 33] able to approach this problem. In [31] the *entropy of a given ensemble* as the normalized logarithm of the number of networks in the ensemble has been introduced. This quantity can be used to asses the role that a given structural characteristics have in shaping the network. In fact, given a real network, a subsequent series of randomized networks ensembles can be build each subsequent ensemble sharing one additional structural characteristic with the given network. The entropy of these subsequent networks ensembles would decreases as we proceed adding constraints and the difference between the entropies in two subsequent ensembles quantifies how restrictive is the introduced additional constraint. In the first part of this paper construct a general statistical mechanics framework for the construction of generalized random network ensembles which satisfy given structural constraints. We call these ensembles "microcanonical". We also describe how to construct "canonical" network ensembles or generalized hidden variable [22, 23, 24, 25, 26, 27] models. Subsequently we make an account of most of the network ensembles that can be formulated: the ensemble of undirected networks with given number of links and nodes, the ensemble of undirected networks with given degree sequence, with given spatial embedding and community structure. Some of these network ensembles were already presented in [31] and we report their derivation here for completeness. This approach is further extended to weighted networks and directed networks. Finally we focus our attention on the structural entropy, i.e. the entropy of an ensemble of uncorrelated undirected simple networks of given degree sequence. The structural entropy of a power-law network with constant average degree is monotonically decreasing as the power-law ex-

ponent  $\gamma \rightarrow 2$ . This result could appear in contradiction with the wide occurrence of power-law degree distribution in complex networks. Here we show by a statistical mechanics model that scale-free degree distribution are the most likely degree distribution at given small value of structural entropy while Poisson degree distributions are the most likely degree distribution of networks with maximal structural entropy.

This result indicates that the scale-free degree distributions emerges naturally when considering networks ensembles with small structural entropy and therefore larger amount of order.

The appearance of the power-law degree distribution reflects the tendency of social, technological and especially biological networks toward "ordering". This tendency is at work regardless of the mechanism which is driving their evolution that can be either a preferential attachment mechanism [3], or a "hidden variables" mechanism [22, 23, 24, 25, 26, 27] or some other statistical mechanics mechanism [20, 21].

## STATISTICAL MECHANICS OF NETWORK ENSEMBLES

A network of  $N$  labeled nodes  $i = 1, 2, \dots, N$  is uniquely defined by its adjacency matrix  $\mathbf{a}$  of matrix elements  $a_{ij} \geq 0$  with  $a_{ij} > 0$  if and only if there is a link between node  $i$  and node  $j$ . Simple networks are networks without tadpoles or double links, i.e.  $a_{ii} = 0$  and  $a_{ij} = 0, 1$ . Weighted networks describe heterogeneous interactions between the nodes and the matrix elements  $a_{ij}$  can take different null or positive values, while directed networks are described by non-symmetric adjacency matrices  $\mathbf{a} \neq \mathbf{a}^T$  where we have indicated by  $\mathbf{a}^T$  the transpose of the matrix  $\mathbf{a}$ .

A structural constraint on a network can always be formulated as a constraint on the adjacency matrix of the graph, i.e.

$$\vec{F}(\mathbf{a}) = \vec{C}. \quad (1)$$

In order to describe "microcanonical" network ensembles with given structural constraints in [31] and in the following we will use a statistical mechanics perspective. Therefore we define a partition function  $Z$  of the ensemble in the following way

$$Z = \sum_{\mathbf{a}} \delta[\vec{F}(\mathbf{a}) - \vec{C}] e^{\sum_{ij} h_{ij} \Theta(a_{ij}) + r_{ij} a_{ij}} \quad (2)$$

where, for simplifying the problem,  $\vec{F}(\mathbf{a})$  and  $a_{ij}$  take only integer values, and  $\delta[\cdot]$  indicate the Kronecker delta and  $\Theta(x) = 1$  if  $x > 0$  and  $\Theta(x) = 0$  if  $x = 0$ . Moreover, in (79), the auxiliary fields  $h_{ij}$  have been introduced as in classical statistical mechanics. The entropy per node

$\Sigma$  of the network ensemble is defined as

$$\Sigma = \frac{1}{N} \ln(Z) \Big|_{h_{ij}=r_{ij}=0 \forall (i,j)}. \quad (3)$$

The marginal probability for a certain value of the element  $a_{ij}$  of the adjacency matrix is given by

$$\pi_{ij}(A) = \frac{1}{Z} \sum_{\mathbf{a}} \delta(a_{ij} - A) \delta(\vec{F}(\mathbf{a}) - \vec{C}). \quad (4)$$

The probability of a link  $p_{ij}$  is given by

$$p_{ij} = \frac{\partial \ln Z}{\partial h_{ij}} \Big|_{h_{ij}=r_{ij}=0 \forall (i,j)}. \quad (5)$$

In an ensemble of weighted network we can define also the average weight  $w_{ij}$  of a link between node  $i$  and node  $j$  as equal to

$$w_{ij} = \frac{\partial \ln Z}{\partial r_{ij}} \Big|_{h_{ij}=r_{ij}=0 \forall (i,j)} \quad (6)$$

In a "microcanonical" network ensemble all the networks that satisfy a given structural constraint have equal probability. Therefore the probability of a network  $G$ , described by the adjacency matrix  $\mathbf{a}$ , is given in the "microcanonical" ensemble by

$$P_M(\mathbf{a}) = e^{-N\Sigma} \delta[\vec{F}(\mathbf{a}) - \vec{C}] \quad (7)$$

If we allow for "soft" structural constraints in network ensemble we can describe "canonical" network ensemble. In a "canonical" network ensemble each network  $\mathbf{a}$  has a different probability given by

$$P_C(\mathbf{a}) = \prod_{ij} \pi_{ij}(a_{ij}) \quad (8)$$

expression that for ensemble of simple networks take the form

$$P_C(\mathbf{a}) = \prod_{ij} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}}. \quad (9)$$

If the link probabilities  $\pi_{ij}(a_{ij})$  are chosen equal to (4) and (5), then we have that the structural constraints  $\vec{F}(\mathbf{a}) = \vec{C}$  are satisfied in average, i.e.

$$\langle \vec{F}(\mathbf{a}) \rangle_{P_C(\mathbf{a})} = \vec{C} \quad (10)$$

where the average  $\langle \cdot \rangle_{P_C(\mathbf{a})}$  indicates the average over the canonical ensembles (8). The statistical mechanics formulation of network ensemble is always well defined. For network structural constraints that do not correspond to feasible networks [34] the entropy of the network ensemble is nevertheless  $\Sigma = -\infty$ . Although the definition of the statistical mechanics problem is always well defined, the calculation of the partition function by saddle point

approximation can only be performed if the number of constraints  $F_\alpha$  with  $\alpha = 1, \dots, M$  is at most extensive, i.e.  $M = \mathcal{O}(N)$ . In addition to that, in the paper we are going to consider only linear constraints on the adjacency matrix. Further developments on this statistical mechanics framework will involve perturbative approach to solve non linear structural constraints.

## UNDIRECTED SIMPLE NETWORKS

In an undirected simple network the adjacency matrix elements are zero/ one ( $a_{ij} = 0, 1$ ) and the tadpoles are forbidden ( $a_{ii} = 0 \forall i$ ). We can consider for these networks different types of structural constraints. In the following we list few of them of particular interest.

- *i)* The ensemble  $G(N, L)$  of random networks with given number of nodes  $N$  and links  $L = \sum_{i < j} a_{ij}$  (providing in this way a statistical mechanics formulation of the  $G(N, L)$  random ensemble). In this case we have the structural constraint

$$\vec{F}(\mathbf{a}) - \vec{C} = \sum_{i < j} a_{ij} - L = 0 \quad (11)$$

- *ii)* The configuration model, i.e. the ensemble of networks with given degree sequence  $\{k_1, \dots, k_N\}$  with  $k_i = \sum_j a_{ij}$ . In this case the structural constraints are given by

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_j a_{\alpha j} - k_\alpha = 0 \quad (12)$$

for  $\alpha = 1, \dots, N$ .

- *iii)* The network with given degree sequence  $\{k_1, \dots, k_N\}$  and given average nearest neighbor connectivity  $k_{nn}(k) = [\sum_{i,j} \delta(k_i - k) a_{ij} k_j] / (k N_k)$  of nodes of degree  $k$  (with  $N_k$  indicating the number of nodes of degree  $k$  in the network). In this case the structural constraints are given by

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_j a_{\alpha j} - k_\alpha = 0 \quad (13)$$

for  $\alpha = 1, \dots, N$  and

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_{ij} \delta(k_i - k) a_{ij} k_j - k N_k k_{nn}(k) \quad (14)$$

for  $\alpha = N + 1, \dots, N + K$ . with  $K$  indicating the maximal connectivity of the network.

- *iv)* The network ensemble with given degree sequence and given community structure. For these network we assume that each node is assigned a feature  $\{q_1, \dots, q_N\}$  and we fix the number of links between nodes of different features  $A(q, q') =$

$\sum_{i < j} \delta(q_{ij} - q) \delta(\overline{q}_{ij} - q') a_{ij}$  with  $q_{ij} = \min(q_i, q_j)$  and  $\overline{q}_{ij} = \max(q_i, q_j)$ . In this case the structural constraints are given by

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_j a_{\alpha j} - k_\alpha = 0 \quad (15)$$

for  $\alpha = 1, \dots, N$  and

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_{i < j} \delta(q_{ij} - q) \delta(\overline{q}_{ij} - q') a_{ij} - A(q, q') \quad (16)$$

for  $\alpha = N + 1, \dots, N + Q(Q + 1)/2$  with  $Q$  equal to the number of different features of the nodes. Here an in the following in order to have an extensive number of constraints we assume  $Q = \mathcal{O}(N^{1/2})$ .

- *v)* The ensemble of networks with given degree sequence and dependence of the link probability on the distance of the nodes in an embedding geometrical space. In this ensemble we consider fixed spatial distribution of nodes in space  $\{\vec{r}_1, \dots, \vec{r}_N\}$  and we consider all the networks compatible with the given degree sequence and the number of links linking nodes in a given distance interval. Therefore we take  $\Lambda$  distance intervals  $I_\ell = [d_\ell, d_\ell + (\Delta d)_\ell]$  with  $\ell = 1, \dots, \Lambda$ , and we fix the number of links linking nodes in a given distance interval. The structural constraint involved therefore the vector  $B(d_\ell) = \sum_{i < j} \chi_\ell(d_{ij}) a_{ij}$  where  $d_{ij} = d(\vec{r}_i, \vec{r}_j)$  is the distance between node  $i$  and  $j$  in the embedding space and the characteristic function  $\chi_\ell(x) = 1$  if  $x \in [d_\ell, d_\ell + (\Delta d)_\ell]$  and  $\chi_d(x) = 0$  otherwise. In this case the structural constraints can be expressed as

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_j a_{\alpha j} - k_\alpha = 0 \quad (17)$$

for  $\alpha = 1, \dots, N$  and

$$F_\alpha(\mathbf{a}) - C_\alpha = \sum_{i < j} \chi_\ell(d_{ij}) a_{ij} - B(d_\ell) \quad (18)$$

for  $\alpha = N + 1, \dots, N + \Lambda$ .

## The $G(N, L)$ and the $G(N, p)$ ensembles

The networks in the  $G(N, L)$  ensemble have given number of nodes  $N$  and links  $L$ . The entropy of this ensemble is given by the logarithm of the binomial

$$N \Sigma_0 = \left( \frac{N(N-1)}{2} \right) \quad (19)$$

(we always assume distinguishable nodes in the networks [29]). The probability  $p_{ij}$  of a given link  $(i, j)$  is given by  $p_{ij}^{(0)} = L/(N(N-1)/2)$  for every couple of nodes  $i, j$ . The ensemble  $G(N, p)$  is the "canonical" ensemble corresponding to the "microcanonical"  $G(N, L)$  ensemble.

### The configuration ensemble

In the configuration ensemble we consider all the networks with given degree sequence. Using (2) (12) the partition function of the ensemble can be explicitly written as

$$Z_1 = \sum_{\{a_{ij}\}} \prod_i \delta(k_i - \sum_j a_{ij}) e^{\sum_{i<j} h_{ij} a_{ij}} \quad (20)$$

Expressing the delta's in the integral form with Lagrangian multipliers  $\omega_i$  for every  $i = 1, \dots, N$  we get

$$Z_1 = \int \mathcal{D}\omega e^{-\sum_i \omega_i k_i} \prod_{i<j} (1 + e^{\omega_i + \omega_j + h_{ij}}) \quad (21)$$

where  $\mathcal{D}\omega = \prod_i d\omega_i / (2\pi)$ . We solve this integral by saddle point equations accounting also for second order terms of the expansion. The entropy of this ensemble of networks can be approximated in the large network limit  $N \gg 1$  with

$$\begin{aligned} N\Sigma_1^{und} &\simeq - \sum_i \omega_i^* k_i + \sum_{i<j} \ln(1 + e^{\omega_i^* + \omega_j^*}) \\ &\quad - \frac{1}{2} \sum_i \ln(2\pi\alpha_i) \end{aligned} \quad (22)$$

with the Lagrangian multipliers  $\omega_i$  satisfying the saddle point equations

$$k_i = \sum_{j \neq i} \frac{e^{\omega_i^* + \omega_j^*}}{1 + e^{\omega_i^* + \omega_j^*}}, \quad (23)$$

and the coefficients  $\alpha_i$  defined as

$$\alpha_i \simeq \sum_j \frac{e^{\omega_i^* + \omega_j^*}}{(1 + e^{\omega_i^* + \omega_j^*})^2}, \quad (24)$$

The probability of a link  $i, j$  in this ensemble is given by

$$p_{ij}^{(1)} = \frac{e^{\omega_i^* + \omega_j^*}}{1 + e^{\omega_i^* + \omega_j^*}}. \quad (25)$$

In particular in this ensemble  $p_{ij} \neq f(\omega_i)f(\omega_j)$ , consequently *the model retains some "natural" correlations* [27] given by the degree sequence and the constraint that we consider only simple networks. In fact these are nothing else than the correlations of the configuration model [35].

The "canonical model" corresponding to the configuration model is then a "hidden variable" models where each node  $i$  is assigned a "hidden variable"  $\omega_i$  and the probability for each link follow (25). Similar expressions were already derived in different papers [22, 25, 27] but with a different interpretation. Here the "hidden variables"  $\omega_i$  are simply fixing the average degrees of each node. We note here that the derivation of [26] guarantees that

in the "canonical" model the connectivity of each node is distributed according to a Poisson distribution with average  $\sum_j p_{ij}$ .

The form of the probability  $p_{ij}$  is such that when inferring the values of the "hidden variables"  $\omega_i$  for a "canonical" network in this ensemble by maximum likelihood methods, we obtain the  $\omega'_i = \omega_i$  in the large network limit [27].

### Uncorrelated networks

The case in which there is a structural cutoff in the network  $k_i < \sqrt{\langle k \rangle N}$  is of particular interest. In this case we can approximate Eq. (23) by  $e^{\omega_i^*} \simeq k_i / \sqrt{\langle k \rangle N}$ ,  $\alpha_i \simeq k_i$ . In this limit the network is *uncorrelated* the probabilities of a link are given by  $p_{ij}^{(1), uncorr} = k_i k_j / (\langle k \rangle N)$ , since the  $\omega_i^* < 0$ . We call the entropy of these uncorrelated ensembles the *structural entropy*  $\Sigma_S$  and we can evaluate it providing the explicit expression

$$\begin{aligned} N\Sigma_S &\simeq - \sum_i \ln[k_i / \sqrt{\langle k \rangle N}] k_i - \frac{1}{2} \sum_i \ln(2\pi k_i) \\ &\quad + \frac{1}{2} \sum_{ij} \frac{k_i k_j}{\langle k \rangle N} - \sum_{ij} \frac{1}{4} \frac{k_i^2 k_j^2}{(\langle k \rangle N)^2} + \dots \\ &= - \sum_i (\ln k_i - 1) k_i - \frac{1}{2} \sum_i \ln(2\pi k_i) + \\ &\quad \frac{1}{2} \langle k \rangle N [\ln(\langle k \rangle N) - 1] - \frac{1}{4} \left( \frac{\langle k^2 \rangle}{\langle k \rangle} \right)^2 + \dots \end{aligned} \quad (26)$$

Expression (26) gives for the number of networks in the ensemble

$$\mathcal{N}_S^{uncorr} \simeq \frac{(\langle k \rangle N)!!}{\prod_i k_i!} \exp \left[ -\frac{1}{4} \left( \frac{\langle k^2 \rangle}{\langle k \rangle} \right)^2 + \mathcal{O}(\ln N) \right]. \quad (27)$$

From combinatorial arguments we can derive an expression  $\mathcal{N}_C^{uncorr}$  for the number of uncorrelated networks with a given degree sequence which agrees with the above estimate (27) in the limit  $N \gg 1$ , i.e.

$$\ln \mathcal{N}_C^{uncorr} = \ln \mathcal{N}_S^{uncorr} + \mathcal{O}(\ln N). \quad (28)$$

In fact by combinatorial arguments we can show that the number of networks with given degree sequence is given by the following expression in the large  $N$  limit, i.e.

$$\mathcal{N}_C^{uncorr} \propto \frac{(2L-1)!! e^{-\frac{1}{4} \left( \frac{\langle k^2 \rangle}{\langle k \rangle} \right)^2}}{\prod_i k_i!} \quad (29)$$

The factor  $(2L-1)!!$  accounts for the total number of wiring's of the links. In fact if we want to construct a network, given a certain distribution of half-edges through

the  $N$  nodes of the network, as a first step we take a half-edge and we match it with one of the  $2L - 1$  other half-edge of the network. Secondly we match a new half-edge with one of the  $2L - 3$  remaining half-edges. Repeating this procedure we get one out of  $(2L - 1)!!$  possible wiring of the links. This number includes also the wiring of the links which gives rise to networks with double links. To estimate the number of such undesired wiring we assume that the network is random, i.e. that the probability that a node with  $k_i$  half-edges connects to a node with  $k_j$  half-edges is a Poisson variable with average  $k_i k_j / (\langle k \rangle N)$ . In this hypothesis the probability  $\Pi$  that the network does not contain double links is equal to [36]

$$\Pi = \prod_{i < j} \left( 1 + \frac{k_i k_j}{\langle k \rangle N} \right) e^{-\frac{k_i k_j}{\langle k \rangle N}} \sim e^{-\frac{1}{4} \left( \frac{\langle k^2 \rangle}{\langle k \rangle} \right)^2}. \quad (30)$$

Finally in the expression (29) for  $\mathcal{N}_c$  there is an additional term which takes into account the number of wiring of the links giving rise to equivalent networks without double links. This term is given by the number of possible permutation of the half-edges at each node, i.e.  $\prod_i k_i!$ . We note here that a similar result was derived by mathematicians for the case in which the maximal connectivity  $K < N^{1/3}$  [37] and an inequality was proved for the case  $K > N^{1/3}$  [38]. Now we extend these results by statistical mechanics methods to uncorrelated networks with maximal connectivity  $K < \sqrt{\langle k \rangle N}$ .

### The entropy of a network ensemble with fixed degree correlations

We consider now network ensembles with given degree correlations and given average degree of neighboring nodes, satisfy the constraints defined in Eqs. (13) and (14). We can proceed to the evaluation of the probability of a link  $p_{ij}^{(2)}$  and the calculation of the entropy of the ensemble as in the configuration model. In this case we have to introduce the Lagrangian multipliers  $\omega_i$  fixing the degree of node  $i$  and the Lagrangian multipliers  $A_k$  fixing the average degree of nodes of degree  $k$ .

The partition function of this ensemble can be evaluated at the saddle point giving for the entropy of the ensemble, in the thermodynamic limit value

$$\begin{aligned} N\Sigma_2^{und} \simeq & - \sum_i \omega_i^* k_i - \sum_k A_k^* k_{nn}(k) k N_k \\ & + \sum_{i < j} \ln(1 + e^{\omega_i^* + \omega_j^* + k_i A_{k_i}^* + k_j A_{k_j}^*}) \\ & - \frac{1}{2} \sum_i \ln(2\pi\alpha_i) - \frac{1}{2} \sum_k \ln(2\pi\alpha_k) \end{aligned} \quad (31)$$

where  $\omega_i^*$  and  $A_k^{star}$  satisfy the saddle point equations

$$\begin{aligned} k_i &= \sum_{j \neq i} \frac{e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}{1 + e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}, \\ k_{nn}(k) &= \frac{1}{k N_k} \sum_i \delta(k_i - k) \sum_{j \neq i} k_j \frac{e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}{1 + e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}. \end{aligned} \quad (32)$$

and where with  $\alpha_i, \alpha_k$  are approximately equal to the following expressions

$$\begin{aligned} \alpha_i &\simeq \sum_j \frac{e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}{\left( 1 + e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*} \right)^2} \\ \alpha_k &\simeq \sum_i \delta(k_i - k) \sum_{j \neq i} k_j^2 \frac{e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}{\left( 1 + e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*} \right)^2}. \end{aligned} \quad (33)$$

The probability  $p_{ij}^{(2)}$  of the link  $(i, j)$  in this ensemble is given by

$$p_{ij}^{(2)} = \frac{e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}{1 + e^{\omega_i^* + \omega_j^* + k_j A_{k_i}^* + k_i A_{k_j}^*}}. \quad (34)$$

This formula generalize the "hidden variable" formula of the configuration model to networks with strong degree-degree correlations. In particular in order to build a "canonical" network with strong degree degree correlation we can consider nodes with "hidden variables"  $\theta_i$  and  $G_\theta$  and a probability  $p_{ij}$  to have a link between a node  $i$  and a node  $j$  given by

$$p_{ij} = \frac{\theta_i \theta_j (G_{\theta_i})^{\theta_j} (G_{\theta_j})^{\theta_i}}{1 + \theta_i \theta_j (G_{\theta_i})^{\theta_j} (G_{\theta_j})^{\theta_i}}. \quad (35)$$

### The entropy of network ensemble with given degree sequence and given community structure

The partition function of network ensembles with given degree sequence (15) and given community structure (16) can be evaluated by saddle point approximation in the large network limit as long as  $Q = \mathcal{O}(N^{1/2})$ .

Following the same steps as in the previous case we find that the entropy for such an ensemble is given by

$$\begin{aligned} N\Sigma_c \simeq & - \sum_i k_i \omega_i^* - \sum_{q \leq q'} A(q, q') w(q, q')^* \\ & + \sum_{i < j} \ln \left( 1 + e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})} \right) \\ & - \frac{1}{2} \sum_i \ln(2\pi\alpha_i) - \frac{1}{2} \sum_{q < q'} \ln(2\pi\alpha_{q, q'}) \end{aligned} \quad (36)$$

with the Lagrangian multipliers  $\{\omega_i^{star}\}, \{w_{q, q'}^{star}\}$  satisfy-

ing the saddle point equations

$$k_i = \sum_{j \neq i} \frac{e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})}}{1 + e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})}} \quad (37)$$

$$A(q, q') = \sum_{i < j} \delta(q_{ij} - q) \delta(\bar{q}_{ij} - q') \times \frac{e^{\omega_i^* + \omega_j^* + w^*(q, q')}}{1 + e^{\omega_i^* + \omega_j^* + w^*(q, q')}},$$

and with  $\alpha_i, \alpha_{q, q'}$  that can be approximated by

$$\alpha_i \simeq \sum_j \frac{e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})}}{\left(1 + e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})}\right)^2} \quad (38)$$

$$\alpha_{q, q'} \simeq \sum_{i < j} \delta(q_{ij} - q) \delta(\bar{q}_{ij} - q') \times \frac{e^{\omega_i^* + \omega_j^* + w^*(q, q')}}{\left(1 + e^{\omega_i^* + \omega_j^* + w^*(q, q')}\right)^2}.$$

In this ensemble the probability for a link  $p_{ij}^{(c)}$  between a node  $i$  and a node  $j$  is equal to

$$p_{ij}^{(c)} = \frac{e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})}}{1 + e^{\omega_i^* + \omega_j^* + w^*(q_{ij}, \bar{q}_{ij})}}. \quad (39)$$

Assigning each node a "hidden variable"  $\theta_i$  and to each pair of communities the symmetric matrix  $V(q, q')$  we can construct the "hidden variable" or "canonical" ensemble by extracting each link with probability

$$p_{ij} = \frac{\theta_i \theta_j V(q_i, q_j)}{1 + \theta_i \theta_j V(q_i, q_j)} \quad (40)$$

### The entropy of a network ensemble with given distance between the nodes

Finally we consider the ensemble of undirected networks living in a generic embedding space and with structural constraints described by (17) and (18). Following the same steps as in the previous cases we find that the entropy for such an ensemble in the large network limit is given by

$$N\Sigma_d \simeq - \sum_i k_i \omega_i^* - \sum_{\ell=1}^{\Lambda} B(d_\ell) g(d_\ell)^* + \sum_{i < j} \ln \left( 1 + e^{\omega_i^* + \omega_j^* + \sum_{\ell} \chi_{\ell}(d_{ij}) g^*(d_\ell)} \right) - \frac{1}{2} \sum_i \ln(2\pi \alpha_i) - \frac{1}{2} \sum_{\ell=1}^{\Lambda} \ln(2\pi \alpha_\ell) \quad (41)$$

with the Lagrangian multipliers  $\{\omega_i\}, \{g_d\}$  satisfying the saddle point equations

$$k_i = \sum_{j \neq i} \frac{e^{\omega_i^* + \omega_j^* + \sum_{\ell} \chi_{\ell}(d_{ij}) g^*(d_\ell)}}{1 + e^{\omega_i^* + \omega_j^* + \sum_{\ell} \chi_{\ell}(d_{ij}) g^*(d_\ell)}} \quad (42)$$

$$B(d_\ell) = \sum_{i < j} \chi_{\ell}(d_{ij}) \frac{e^{\omega_i^* + \omega_j^* + g^*(d_\ell)}}{1 + e^{\omega_i^* + \omega_j^* + g^*(d_\ell)}},$$

and the variables  $\alpha_i, \alpha_{q, q'}$  approximated by the expressions

$$\alpha_i \simeq \sum_j \frac{e^{\omega_i^* + \omega_j^* + \sum_{\ell} \chi_{\ell}(d_{ij}) g^*(d_\ell)}}{\left(1 + e^{\omega_i^* + \omega_j^* + \sum_{\ell} \chi_{\ell}(d_{ij}) g^*(d_\ell)}\right)^2}$$

$$\alpha_\ell \simeq \sum_{i, j} \chi_{\ell}(d_{ij}) \frac{e^{\omega_i^* + \omega_j^* + g^*(d_\ell)}}{\left(1 + e^{\omega_i^* + \omega_j^* + g^*(d_\ell)}\right)^2} \quad (43)$$

The probability for a link between node  $i$  and  $j$  is equal to

$$p_{ij}^{(d)} = \sum_{\ell} \chi_{\ell}(d_{ij}) \frac{e^{\omega_i^* + \omega_j^* + g^*(d_\ell)}}{1 + e^{\omega_i^* + \omega_j^* + g^*(d_\ell)}}. \quad (44)$$

Therefore the "hidden variable" model associated to this ensemble correspond to a model where we fix the "hidden variables"  $\theta_i$  and  $W(d_\ell)$  and we draw a link between node  $i$  and node  $j$  according to

$$p_{ij} = \sum_{\ell} \chi_{\ell}(d_{ij}) \frac{\theta_i \theta_j W(d_\ell)}{1 + \theta_i \theta_j W(d_\ell)}. \quad (45)$$

## WEIGHTED NETWORKS

Many networks not only have a non trivial topological structure but are also characterized by weighted links. We will assume in this paper that the weight of a link can assume only integer values and consequently a link between a node  $i$  and node  $j$  is characterized by an integer number  $a_{ij} \geq 1$ , this is not a very stringent constraints since we can assume to have always finite networks (studied in the thermodynamic limit). In a weighted network the degree and the strength  $s_i$  of the node  $i$  are defined as

$$k_i = \sum_{j \neq i} \Theta(a_{ij})$$

$$s_i = \sum_{j \neq i} a_{ij} \quad (46)$$

where  $\Theta(x) = 0$  if  $x = 0$  and  $\Theta(x) = 1$  if  $x > 0$ . It is possible to define series of weighted networks by considering networks with fixed total strength, with given strength sequence, with given strength and degree sequence and proceeding by adding additional features as in the unweighted case. Here in the following we study the most relevant cases:

- We first consider the network ensemble with given total strength  $S$ . The structural constraint in this case is equal to

$$F(\mathbf{a}) - C = \sum_{i < j} a_{ij} - S = 0. \quad (47)$$

- ii) We consider the network with given strength sequence  $s_1, \dots, s_N$ . The structural constraints are for this ensemble given by

$$F(\mathbf{a})_\alpha - C_\alpha = \sum_j a_{\alpha j} - s_\alpha = 0. \quad (48)$$

for  $\alpha = 1, \dots, N$ .

- iii) Finally we consider the network ensemble with given degree sequence  $\{k_1, \dots, k_n\}$  and strength sequence  $\{s_1, \dots, s_N\}$ . For this ensemble the structural constraints are given by

$$F(\mathbf{a})_\alpha - C_\alpha = \sum_j \Theta(a_{\alpha j}) - k_\alpha = 0. \quad (49)$$

for  $\alpha = 1, \dots, N$  and

$$F(\mathbf{a})_\alpha - C_\alpha = \sum_j a_{\alpha j} - s_\alpha = 0. \quad (50)$$

for  $\alpha = 1, \dots, 2N$ .

#### The entropy of weighted network ensembles with given total strength $S$

The entropy of this ensemble is given by

$$N\Sigma_1^W = \ln \left[ \left( \frac{\frac{N(N-1)}{2} + S}{\frac{N(N-1)}{2}} \right) \right]$$

The average value of a the weight of the link from  $i$  to  $j$  is given by

$$w_{ij} = \langle a_{ij} \rangle_1^W = \frac{S}{\frac{N(N-1)}{2}} \quad (51)$$

and the probability of a link between node  $i$  and  $j$  is equal to

$$p_{ij}^{W,1} = \frac{S}{S + \frac{N(N-1)}{2}}. \quad (52)$$

Therefore the simple networks with adjacency matrix  $((A_{ij}))$  that can be constructed from the weighed networks with adjacency matrix  $((a_{ij}))$  by putting  $A_{ij} = \Theta(a_{ij}) \forall i, j$  is uncorrelated. The canonical ensemble is given by Eq. (8) with

$$\pi_{ij}(a_{ij}) = \frac{e^{\omega a_{ij}}}{1 - e^\omega} \quad (53)$$

and  $\omega = -\ln[1 + N(N-1)/(2S)]$ .

#### The entropy of weighted network ensembles with given strength sequence

To calculate the entropy of undirected networks with a given strength sequence of degrees  $\{s_i\}$  we proceed by the saddle point approximation as in previous cases We find that the entropy of this ensemble of networks is given by

$$N\Sigma_1^W \simeq - \sum_i \omega_i^* s_i - \sum_{i < j} \ln(1 - e^{\omega_i^* + \omega_j^*}) - \frac{1}{2} \sum_i \ln(2\pi \lambda_i) \quad (54)$$

with the Lagrangian multipliers  $\omega_i^*$  satisfying the saddle point equations

$$s_i = \sum_{j \neq i} \frac{e^{\omega_i^* + \omega_j^*}}{1 - e^{\omega_i^* + \omega_j^*}}. \quad (55)$$

and with  $\lambda_i$  being the eigenvectors of the Jacobian of the function

$$\mathcal{F} = \sum_{i < j} \ln [1 - e^{-\omega_i - \omega_j}]. \quad (56)$$

The average value of a the weight of the link from  $i$  to  $j$  is given by

$$\langle a_{ij} \rangle_1^W = \frac{e^{\omega_i^* + \omega_j^*}}{1 - e^{\omega_i^* + \omega_j^*}}. \quad (57)$$

and the probability of a link between node  $i$  and  $j$  is equal to

$$p_{ij}^{W,1} = e^{\omega_i^* + \omega_j^*}. \quad (58)$$

Therefore as it has been observed in [?] only by rewiring the links of a network allowing for multilinks we get a network structure which is uncorrelated.

The canonical ensemble (8) in this case can be constructed by assigning to every possible link  $(i, j)$  the weight  $a_{ij}$  with the probability

$$\pi_{ij}(a_{ij}) = \frac{e^{(\omega_i^* + \omega_j^*) a_{ij}}}{1 - e^{\omega_i^* + \omega_j^*}}. \quad (59)$$

#### The entropy of weighted network ensembles with given strength /degree sequence

The entropy of weighted networks with a given strength and degree sequence  $\{s_i, k_i\}$  in the large size network limit is given by

$$\begin{aligned} N\Sigma_2^W = & - \sum_i \omega_i^* s_i - \sum_i \psi_i^* k_i - \sum_i \\ & + \sum_{i < j} \ln \left[ 1 + e^{\psi_i^* + \psi_j^*} \frac{1}{e^{-\omega_i^* - \omega_j^*} - 1} \right] \\ & + \frac{1}{2} \sum_{\ell=1} 2N \sum_i \ln(2\pi \lambda_\ell) \end{aligned} \quad (60)$$

with the Lagrangian multipliers satisfying the saddle point equations

$$k_i = \sum_{j \neq i} \frac{e^{\psi_i^* + \psi_j^*}}{e^{\psi_i^* + \psi_j^*} + e^{-(\omega_i^* + \omega_j^*)} - 1},$$

$$s_i = \sum_{j \neq i} \frac{e^{-(\omega_i^* + \omega_j^*) + (\psi_i^* + \psi_j^*)}}{(e^{\psi_i^* + \psi_j^*} + e^{-(\omega_i^* + \omega_j^*)} - 1)(e^{-\omega_i^* - \omega_j^*} - 1)} \quad (61)$$

$$(62)$$

and with  $\lambda_\ell$  being the eigenvectors of the Jacobian of the function

$$\mathcal{F} = \sum_{i < j} \ln \left[ 1 + e^{\psi_i + \psi_j} \frac{1}{e^{-\omega_i - \omega_j} - 1} \right] \quad (63)$$

calculated at the values  $\{\omega_i^*, \psi_i^*\}$ . The average weight of the link  $(ij)$  is given by

$$\langle a_{ij} \rangle_2^W = \frac{e^{-(\omega_i^* + \omega_j^*) + (\psi_i^* + \psi_j^*)}}{(e^{\psi_i^* + \psi_j^*} + e^{-(\omega_i^* + \omega_j^*)} - 1)(e^{-\omega_i^* - \omega_j^*} - 1)} \quad (64)$$

and the probability of a link between node  $i$  and  $j$  is equal to

$$p_{ij}^{W,2} = \frac{e^{\psi_i^* + \psi_j^*}}{e^{\psi_i^* + \psi_j^*} + e^{-(\omega_i^* + \omega_j^*)} - 1} \quad (65)$$

The canonical ensemble (8) in this case can be constructed by assigning to every possible link  $(i, j)$  the weight  $a_{ij}$  with the probability

$$\pi_{ij}(a_{ij}) = \frac{e^{(\psi_i^* + \psi_j^*)\Theta(a_{ij})} e^{(\omega_i^* + \omega_j^*)a_{ij}}}{1 + e^{\psi_i^* + \psi_j^*} \frac{1}{e^{-\omega_i^* - \omega_j^*} - 1}}. \quad (66)$$

## DIRECTED NETWORKS

An undirected network is determined by a symmetric adjacency matrix, while the matrix of a directed network is in general non-symmetric. Consequently the degrees of freedom of a directed network are more than the degrees of freedom of an undirected network. In the following we only consider the network ensemble with

- *i)* Total number of directed links The structural constraint in this case is equal to

$$F(\mathbf{a}) - C = \sum_{ij} a_{ij} - S = 0. \quad (67)$$

- *ii)* Given directed degree sequence  $\{k_1^{in}, k_1^{out}, \dots, k_N^{in}, k_N^{out}\}$ . The structural constraints in this case are

$$F(\mathbf{a})_\alpha - C_\alpha = \sum_j a_{\alpha j} - k_\alpha^{out} = 0. \quad (68)$$

for  $\alpha = 1, \dots, N$  and

$$F(\mathbf{a})_\alpha - C_\alpha = \sum_j a_{j\alpha} - k_\alpha^{in} = 0. \quad (69)$$

for  $\alpha = N + 1, \dots, 2N$ .

### The entropy of directed network ensembles with fixed number of directed links

If we consider the number of directed networks  $\mathcal{N}_0^{dir}$  with given number of nodes and of directed links we find

$$\mathcal{N}_0^{dir} = \binom{N(N-1)}{L^{dir}}. \quad (70)$$

In this case the probability of a directed link is given by

$$p_{ij} = \frac{L}{N(N-1)}. \quad (71)$$

### The entropy of directed network ensembles with given degree sequence

To calculate the entropy of directed networks with a given degree sequence of in/out degrees  $\{k_i^{out}, k_i^{in}\}$  we just have to impose the constraints on the incoming and outgoing connectivity,

$$Z_1^{dir} = \sum_{\{a_{ij}\}} \prod_i \delta(k_i^{(out)} - \sum_j a_{ij}) \prod_i \delta(k_i^{(in)} - \sum_j a_{ji}) \exp\left[\sum_{ij} h_{i,j} a_{ij}\right] \quad (72)$$

Following the same approach as for the undirected case, we find that the entropy of this ensemble of networks is given by

$$\begin{aligned} N\Sigma_1^{dir} \simeq & - \sum_i \omega_i^* k_i^{(out)} - \sum_i k_i^{(in)} \hat{\omega}_i^* \\ & + \sum_{i \neq j} \ln(1 + e^{\omega_i^* + \hat{\omega}_j^*}) \\ & - \frac{1}{2} \sum_i \ln((2\pi)^2 \alpha_i^{(in)} \alpha_i^{(out)}) \end{aligned} \quad (73)$$

with the Lagrangian multipliers satisfying the saddle point equations

$$k_i^{(out)} = \sum_{j \neq i} \frac{e^{\omega_i^* + \hat{\omega}_j^*}}{1 + e^{\omega_i^* + \hat{\omega}_j^*}},$$

$$k_i^{(in)} = \sum_{j \neq i} \frac{e^{\omega_j^* + \hat{\omega}_i^*}}{1 + e^{\omega_j^* + \hat{\omega}_i^*}}. \quad (74)$$



with

$$\begin{aligned}\alpha_i^{(out)} &\simeq \sum_{j \neq i} \frac{e^{\omega_i^* + \hat{\omega}_j^*}}{(1 + e^{\omega_i^* + \hat{\omega}_j^*})^2} \\ \alpha_i^{(in)} &\simeq \sum_{j \neq i} \frac{e^{\omega_j^* + \hat{\omega}_i^*}}{(1 + e^{\omega_j^* + \hat{\omega}_i^*})^2}\end{aligned}\quad (75)$$

The probability for a directed link from  $i$  to  $j$  is given by

$$p_{ij}^{(1,dir)} = \frac{e^{\omega_i^* + \hat{\omega}_j^*}}{1 + e^{\omega_i^* + \hat{\omega}_j^*}}. \quad (76)$$

If the  $\omega_i + \hat{\omega}_j < 0 \forall i, j = 1, \dots, N$  the directed network becomes uncorrelated and we have  $p_{ij}^{1,(dir)} = k_i^{(out)} k_j^{(in)} / \sqrt{\langle k_{in} \rangle N}$ . Given this solution the condition for having uncorrelated directed networks is that the maximal in-degree  $K^{(in)}$  and the maximal out-degree  $K^{(out)}$  should satisfy,  $K^{(in)} K^{(out)} / \sqrt{\langle k_{in} \rangle N} < 1$ . The entropy of the directed uncorrelated network is then given by

$$\begin{aligned}N \Sigma_{1,dir}^{uncorr} &\simeq \ln(\langle k_{in} \rangle N)! - \sum_i \ln(k_i^{(in)}! k_i^{(out)}!) \\ &\quad - \frac{1}{2} \frac{\langle k_{in}^2 \rangle \langle k_{out}^2 \rangle}{\langle k_{in} \rangle \langle k_{out} \rangle}\end{aligned}\quad (77)$$

which has a clear combinatorial interpretation as it happens also for the undirected case.

### NATURAL DEGREE DISTRIBUTION CORRESPONDING TO A GIVEN STRUCTURAL ENTROPY

For power-law networks with power-law exponent  $\gamma \in (2, 3)$  the entropy of the networks with fixed degree sequence  $\Sigma_1$  given by Eq. (22) decreases with the value of the power-law exponent  $\gamma$  when we compare network ensemble with the same average degree [31]. Therefore scale-free networks have much smaller entropy than homogeneous networks. This fact seems to be in contrast with the fact that scale-free networks are the underlying structure of a large class of complex systems. The apparent paradox can be easily be resolved if we consider that many networks are the result of a non-equilibrium

dynamics. Therefore they do not have to satisfy the maximum entropy principle. Nevertheless, in order to give more insight and comment on the universal occurrence of power-law networks in this section we derive the most likely degree distribution of given *structural entropy* when the total number of nodes and links are kept fixed. By structural entropy we define the entropy  $\Sigma_S$  (26) of uncorrelated networks with fixed degree distribution. In order to do that we construct a statistical model very closely related to the urn or “ball in the box” models [21, 39]

We consider degree distributions  $\{N_k\} = \sum_i \delta(k - k_i)$  which arise from the random distribution of the  $2L$  half-edges through the  $N$  nodes of the network. The number of ways  $\mathcal{N}_{\{N_k\}}$  in which we can distribute the  $(2L)$  half-edges in order to have a  $\{N_k\}$  degree distribution are

$$\mathcal{N}_{\{N_k\}} = \frac{(2L)!}{\prod_k (k N_k)!}. \quad (78)$$

We want to find the most likely degree distribution that corresponds to a given value of the structural entropy.

Proceeding as in standard statistical mechanics, we define a normalized partition function  $\mathcal{Z}$  as

$$\mathcal{Z} = \frac{1}{C} \sum_{\{N_k\}} \mathcal{N}_{\{N_k\}} e^{\beta N \Sigma_S(\{N_k\})}. \quad (79)$$

with  $C = (2L)! \exp[\beta(2L)!]$ . The role of the parameter  $\beta$  in Eq. (79) is to fix the average value of the structural entropy  $\Sigma_S$ . When  $\beta \rightarrow \infty$  the structural entropy  $\Sigma_S$  is maximized when  $\beta \rightarrow \beta_{min}$  the structural entropy  $\Sigma_S$  is minimized.

In equation (79) the sum  $\sum'$  over the  $\{N_k\}$  distributions is extended only to  $\{N_k\}$  for which the total number of nodes  $N$  and the total number of links  $L$  in the network is fixed, i.e.

$$\begin{aligned}\sum_k N_k &= N \\ \sum_k k N_k &= 2L.\end{aligned}\quad (80)$$

To enforce these conditions we introduce in (79) the delta functions in the integral form providing the expression

$$\begin{aligned}\mathcal{Z} &= \frac{1}{(2L)!} \int \frac{d\lambda}{2\pi} \int dS \int \frac{d\mu}{2\pi} \int \frac{d\nu}{2\pi} \sum_{\{N_k\}} \exp \left[ -\beta \sum_k N_k \ln k! - \frac{\beta}{4} \left( \frac{S}{\langle k \rangle} \right)^2 - \sum_k \ln[(k N_k)!] \right. \\ &\quad \left. - i\lambda(2L - \sum_k N_k k) - i\mu(N - \sum_k N_k) - i\nu(NS - \sum_k k^2 N_k) \right].\end{aligned}\quad (81)$$

$$\begin{aligned} \mathcal{Z} &= \int dS \int \frac{d\lambda}{2\pi} \int \frac{d\mu}{2\pi} \int \frac{d\nu}{2\pi} \exp \left[ -i\lambda 2L - i\mu N - i\nu NS - \frac{\beta}{4} \left( \frac{S}{\langle k \rangle} \right)^2 + \sum_k \ln G_k(\lambda, \mu, \nu) \right] = \\ &= \int dS \int \frac{d\lambda}{2\pi} \int \frac{d\mu}{2\pi} \int \frac{d\nu}{2\pi} \exp[Nf(\lambda, \mu, \nu, S)] \end{aligned} \quad (82)$$

where

$$G_k(\lambda, \mu, \nu) = \sum_{N_k} \frac{1}{(kN_k)!} \left\{ kN_k \left[ i\lambda + i\frac{\mu}{k} + i\nu k - \frac{\beta}{k} \ln(k!) \right] \right\}. \quad (83)$$

Assuming that the sum over all  $N_k$  can be approximated by the sum over all  $L_k = kN_k = 1, 2, \dots, \infty$  we get  $\ln G_k(\lambda, \mu, \nu) = \exp \left[ i\lambda + i\mu/k - \frac{\beta}{k} \ln(k!) + i\nu k \right]$  and

$$\begin{aligned} f(\lambda, \mu, \nu, S) &= -i\langle k \rangle \lambda - i\mu - i\nu S - \frac{\beta}{4} \left( \frac{S}{\langle k \rangle} \right)^2 \\ &\quad + \frac{1}{N} \sum_k e^{i\lambda + i\mu/k - \frac{\beta}{k} \ln(k!) + i\nu k} \end{aligned} \quad (84)$$

where  $\langle k \rangle = 2L/N$  indicates the average degree of the network. By evaluating (82) at the saddle point, deriving the argument of the exponential respect to  $\lambda$  and  $\nu$ , we obtain

$$\begin{aligned} 1 &= \frac{1}{N} \sum_k \frac{1}{k} e^{i\lambda + i\mu/k - \frac{\beta}{k} \ln(k!) + i\nu k}, \\ \langle k \rangle &= \frac{1}{N} \sum_k e^{i\lambda + i\mu/k - \frac{\beta}{k} \ln(k!) + i\nu k} \\ S &= \frac{1}{N} \sum_k k^2 e^{i\lambda + i\mu/k - \frac{\beta}{k} \ln(k!) + i\nu k} \end{aligned}$$

$$i\nu N = -\beta \frac{S}{2\langle k \rangle^2}. \quad (85)$$

These equations always have a solution for sparse networks with  $L = \mathcal{O}(N)$  provided that  $\beta > 1$  and  $\langle k \rangle > 1$ . The marginal probability that  $L_k = kN_k$  is given by

$$\begin{aligned} P(L_k = kN_k) &= \frac{1}{(kN_k)!} e^{-\beta N_k (\ln(k!) + i\lambda k + i\mu + i\nu k^2)} \\ &\quad \times \frac{\mathcal{Z}_k(L, kN_k, N)}{\mathcal{Z}(L)}, \end{aligned} \quad (86)$$

with

$$\mathcal{Z}_k(L, \ell, N) = \int dS \int \frac{d\lambda}{2\pi} \int \frac{d\mu}{2\pi} \int \frac{d\nu}{2\pi} \exp[Nf_k(\lambda, \mu, \nu, S, \ell)] \quad (87)$$

and

$$\begin{aligned} f_k(\lambda, \mu, \nu, \ell) &= -i(\langle k \rangle - \ell/N)\lambda - i\mu(1 - \ell/(kN)) + \\ &\quad -i\nu(S - k\ell/N) - \frac{\beta}{2} \left( \frac{S^2}{\langle k \rangle} + \right) \frac{1}{N} \\ &\quad + \ln \left[ \sum_{s \neq k} \frac{1}{(sN_s)!} \exp[sN_s(i\lambda + i\mu/s + i\nu s - \frac{\beta}{s} \ln(s!))] \right] \end{aligned} \quad (88)$$

If we develop (86) for  $\ell \ll L$  and we use the Stirling approximation for factorials, we get that each variable  $L_k$  is a Poisson variable with mean  $\langle L_k \rangle$  satisfying

$$\frac{\langle L_k \rangle}{k} = \langle N_k \rangle \simeq k^{-\beta-1} e^{i\lambda + \beta + i\mu/k + i\nu k} \quad (89)$$

where we assume that the minimal connectivity of the network is  $k > 0$ . The average  $\langle N_k \rangle$  is a power-law distribution with a lower and upper effective cutoffs  $-i\mu$  and  $1/(i\nu)$  fixing the average degree  $\langle k \rangle$ , with the Lagrangian parameter  $\lambda$  fixing the normalization constant and finally  $\beta$  fixing the structural entropy. The distribution of  $P(N_k)$  is finally

$$P(N_k) = \frac{k}{(kN_k)!} e^{-\beta N_k \ln(k!) + i\lambda k N_k + i\mu N_k + i\nu (k)^2} \quad (90)$$

In the limit  $\beta \rightarrow \infty$  (90) is extremely peaked around the average degree  $k \simeq k^* = \mathcal{O}(\langle k \rangle)$  of the network and the degree distribution  $N_k$  decays at large value of  $N_k$  as a Poisson distribution, i.e.

$$P(N_k) \simeq \frac{1}{(kN_k)!} e^{kN_k[-\beta \ln(k^*)/k^* + i\lambda k^* + i\mu/k^* + i\nu k^*]}. \quad (91)$$

Therefore for  $\beta \rightarrow \infty$  the network is Poisson like. In the opposite limit of small structural entropy and  $\beta$  small the  $P(N_k)$  distribution (90) develops a fat tail decaying like a power-law (89) with an exponent  $\gamma = \beta + 1$ . Therefore the natural distribution with a small value of the structural entropies are decaying as a power law and smaller values of the power-law exponent correspond to a smaller value of the structural entropy. When the value of the entropy is minimal,  $\beta \rightarrow 1$  the degree distribution (91) has a large tail with an exponent  $\gamma \rightarrow 2$ .

## CONCLUSIONS

In conclusion we have shown that there is a wide set of network ensembles that can be naturally described by statistical mechanics methods. The statistical mechanics method provides the theoretical estimation of the entropy of these ensembles that quantify the cardinality of the network ensembles. We believe that the entropy

of randomized ensembles constructed from a given real networks will be of great applicability for inference problems defined on technological social and biological networks. In this paper we have focused on some theoretical problems that can be approached with the use of this quantity. First we have formulated a series of “canonical” or “hidden variables” models that can be used for generating networks with community structure and spatial embedding. Secondly we have focused on the degree distribution of network. The degree distributions are not all equivalent. In fact the associated structural entropy depends strongly on the distribution. In particular the power-law degree distribution with exponent  $\gamma$  and fixed average degree are associated to a structural entropy that decreases with  $\gamma$ . Nevertheless we have shown that power-law degree distributions are the more likely distributions associated to small structural entropy. This shed light on the evidence that power-law networks constitute a large universality class in complex networks with a non trivial level of organization.

This work was supported by IST STREP GENNETEC contract No. 034952.

- 
- [1] S. N. Dorogovtsev, A. Goltsev and J. F. F. Mendes, arXiv:0705.0010 [cond-mat] (2007).
  - [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D. U. Hwang, Phys. Rep. **424**, 175 (2006).
  - [3] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
  - [4] R. Pastor-Satorras, A. Vázquez and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
  - [5] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
  - [6] J. Berg and M. Lassig, Phys. Rev. Lett. **89**, 228701 (2002).
  - [7] D. J. Watts and S. H. Strogatz, Nature **4**, 393 (1998).
  - [8] E. Ravasz, A. L. Somera, A. D. Mongru, Z. N. Oltvai and A.-L. Barabási, Science **297**, 1551 (2002).
  - [9] S. Carmi, S. Havlin, S. Kirkpatrick, S. Shavitt and E. Shir, PNAS **104**, 11150 (2007).
  - [10] S. N. Dorogovtsev, A. V. Golteev and J. F. F. Mendes, Phys. Rev. Lett. **96**, 040601 (2006).
  - [11] J. I. Alvarez-Hamelin, L. Dall’Asta, A. Barrat and A. Vespignani, cs.Ni/0511007 (2005).
  - [12] M. Girvan and M. E. J. Newman, PNAS **99**, 7821 (2002).
  - [13] L. Danon, A. Díaz-Guilera, J. Duch and A. Arenas, J. Stat Mech. P09008 (2005).
  - [14] M. E. J. Newman and E. A. Leicht, PNAS **104**, 9364 (2007).
  - [15] J. M. Kleiberg, Nature **406**, 846 (2000).
  - [16] S. N. Dorogovtsev, P. L. Krapivsky nad J. F. F. Mendes, Europhys. Lett. **81**, 30004 (2008).
  - [17] M. Boguñá, R. Pastor-Satorras, A. Diaz-Guilera and A. Arenas, Phys. Rev. E **70**, 056122 (2004).
  - [18] M. Barthélemy, R. Pastor-Satorras and A. Vespignani, PNAS **101**, 3747 (2004).
  - [19] G. Bianconi, N. Gulbahce and A. E. Motter, Phys. Rev. Lett. **100**, 118701 (2008).
  - [20] Z. Burda, J. D. Correia and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).
  - [21] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, Nucl. Phys. B **666**, 396 (2003).
  - [22] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066146 (2004).
  - [23] B. Sodeberg, Phys. Rev. E **66**, 066121 (2002).
  - [24] F. Chung and L. Lu, PNAS **100**, 6313 (2002).
  - [25] G. Caldarelli, A. Capocci, P. De Los Rios and M. A. Muñoz Phys. Rev. Lett. **85** 5468 (2002).
  - [26] M. Boguñá and R. Pastor-Satorras Phys. Rev. E **68**, 036112 (2003).
  - [27] D. Garlaschelli and M. I. Loffredo preprint, cond-mat/0609015 (2006).
  - [28] R. F. i Cancho, R. Solé, in *Lecture notes in Physics* (Springer,Berlin 2004).
  - [29] L. Bogacz, Z. Burda and B. Waclaw, Physica A **366**, 587 (2006).
  - [30] M. Bauer, D. Bernard, preprint, arXiv:cond-mat/0206150
  - [31] G. Bianconi, *Europhys. Lett.* **81**, 28005 (2008).
  - [32] G. Bianconi, A. C. C. Coolen and C. J. Perez Vicente (to be published)
  - [33] J. Gomez-Gardenes and V. Latora, preprint arXiv:0712.0278[cond-mat].
  - [34] H. Kim, Z. Toroczkai, I. Miklos, P. L. Erdös and L. A. Székely, unpublished, (2008)
  - [35] M. Molloy and B. A. Reed, Rand. Stru. Alg. **6**, 161 (1995).
  - [36] G. Bianconi, Chaos **17**, 026114 (2007).
  - [37] E. Bender and E. Rodney Canfield, J. Combin. Theory Ser. A **24**, 296 (1978).
  - [38] B. D. McKay, Ars Combin. **19A**, 15 (1985).
  - [39] F. Ritort, *Phys. Rev. Lett.* **75**, 1190 (1995).